Subject : Artificial Intelligence

Instructor : Dr Junaid Akhter

Report : Template Machining through Evolutionary Algorithm

Group Members:

1. Qazi Arsalan Shah
2. Wisha Khursheed
3. Muhamamd Ali Raza

Table Of Contents:

1. Background:

Charles Robert Darwin was an English naturalist, geologist and biologist, known for his contribution to science of evolution.

He was born in Shrews-bury, England in 1809. At 16 he studied medicine at Edinburgh University. Then he changed his subject and started studying to be a clergyman at Christ College, Cambridge. He obtained his degree in Theology in 1831.

From August of 1831 through 1836, he signed on as a naturalist on a scientific voyage aboard the HMS Beagle which sailed the world in an effort to study various aspects of science and the natural world.

He done a remarkable work being a naturalist as presented the his theory of evolution with compelling evidence in his 1859 book “On the Origin of Species”.

Now subsequent topics below cover the overall pipeline followed by scientists.

2. Natural Reality (Phenomenon In Nature):

Darwin’s voyage began on December 1831 and lasted for almost five years. During his journey he kept notes of his keen observation of environment. And collected different specimens. He observed that there was a close relation between different species and there is a diversity in nature. He observed tortoise on different areas and they were different from other tortoise on different area of land. In simple words species were adopted with their environment.

Darwin noticed variations in the beaks and begins to study it. He observed that some finches have long beaks that they use for eating grains while some have short beaks that are best for nut cracking. He claimed that if grains are endangered, then the finches with long beaks wouldn't be able to survive. As a result, they will die or they have moved into other regions. His observation continued for five long years of his voyage, close to nature with a keen eye on phenomena occurring in nature.

Darwin observed the nature closely and tried to find what is happening in nature. He observed that many species very closely related, he deduced that many same kind of species like turtles, mice, birds, beetles change their characteristics (traits) in respond to different conditions in different areas (more specifically in different islands). He also observed relation in between different groups of species as well.

Darwin thought of life as a single **tree, branching and re-branching**. That way, characteristics shared by different living things could be explained by their shared ancestry. He realized that if reproduction went **unchecked**, and grew geometrically, then the world would be overpopulated in a few hundred generations.

3. Theory:

Based on natural phenomenon observed in his voyage and after spending a huge amount of time, Darwin proposed his evolutionary theory in 1859.

In his book “**On the Origin of Species**” he presented his great evolutionary theory.

There is a huge population of species and many different groups of species. It means there is diversity in nature. As in case of tortoise, there were many traits that were not found in other tortoise. And with time many species modified and were completely changed to other species. But that change was not a random change. That change and evolution was to get better and better form of species. Environment wants that, species that are best suited to that environment should remain in it and other are not preferred. In simple words nature is selecting the individuals to survive and to grow their population (But is change and improvement is very slow).

So main point in this theory is of **Natural Selection** i.e. survival of the fittest. In precise words we can write theory as,

“There is a population of individuals (species) in the environment, then nature evaluates all the individuals according to fitness criteria and selects the fittest ones that are well adopted to the environment and then they are allowed to reproduce (selected as parents), off springs by promoting their good qualities to off springs and off springs are not exact replica of their parents. There are some variety operators that maintains diversity in population. Natural selection helps to increase fitness in population and the variety operators maintains diversity in that population.“

The Darwin theory of evolution states that species undergo a process “Natural Selection”. In Natural Selection the heritable traits that are useful for survival in limited resources (less food, predators attack, humidity, etc) are moved to the offspring over many years and the best suitable species remains. In the above example of tortoise, as there was less food in one area so the turtles undergo mutation for the survival so that they have the peak on front that helps them to eat a bit above the ground easily.

Formally, the points of the Theory are:

- The species change over with time, also give rise to new species and all species share a common ancestor.
- Darwin proposed the process “Natural Selection” for the mechanism of evolution. In Natural selection the species under limited resources like less food, environment constraints etc in Nature tend to survive more and give birth to more offspring which take the traits of the parents and regenerate the traits in more frequency.
- In Natural Selection the specie changes its traits over a period of time (generating more child with the more fitter qualities) according to the given environment. To support this point Darwin gave example of the birds. Some birds that eat large seeds tended to have large and strong beak while those birds which ate insects had small and sharp beak. Moreover Natural Selection can be elaborated with the example of “Black and White(tan) Mice ” White mice with fur were easily visible from sky thus can be caught by eagle(, hawk). While on the other hand black mice were not as visible as white. Fur color is heritable Trait that can be passed from one generation to offspring so as the population undergoes crossover, and the next generation have more chances of having a black Fur, and after some steps of generations there will be Higher fraction of population of black mice.
- This Theory is also called “survival of the fittest ”, here fittest means the ability of the specie to survive and reproduce the more suitable (according to the environment) offspring.

4. Extracting Computational Model :

A computational model can be extracted from the Darwin's evolutionary theory which focuses on the process of natural selection. And we can call it as an evolutionary computational model.

Introduction:

As nature of theory suggest (Natural Selection i.e. improvement in next generations), it can be well suited for the problems that involve optimization. As natural selection continuously pushes population towards good quality and improve the individuals that can best suit to the environment. We can construct a computational model based on this idea and can solve problems that need optimization i.e. moves towards optima. We will model a heuristic based based searching approach based on this theory.

In computer science, artificial intelligence, and mathematical optimization, a heuristic (from Greek εὑρίσκω "I find, discover") is a technique designed for solving a problem more quickly when classic methods are too slow, or for finding an approximate solution when classic methods fail to find any exact solution. [wiki]

So we are going to model our evolutionary theory into a computational model and its crux is "Natural Selection Process".

Model:

Computational model is as follow,

We are given a population of individuals, environment then selects best individuals that are well suited for it by process of natural selection also called as survival of fittest. Natural selection is basically improving fitness of population. This fitness is measured according to a fitness function that calculates fitness of selected candidates. Now based on that fitness values, suitable candidates are allowed to become parents and produce next generation by undergoing recombination (crossover) and mutation. Crossover and mutation are variation operators. Crossover is applied to two or more parents by swapping portions of parents and new off springs are created. Mutation is applied only to one candidate by changing few parts of it to make it a bit different from parents (Diversity).

This process is to be repeated until we reach our optimal stage or we have exhausted some number of generations.

It is clear that there is randomness in this model of computation. Selection of parents can be random and there are many ways to select parents but it can be more application specific.

During crossover , which part is to be crossed is random selection. Similarly, which part is to be mutated is random. But great amount of work is done on evolutionary computation and researchers have also mentioned different efficient ways of population selection, crossover methods and smart mutation.

Basic pseudo code / Algorithm of our model is given below,

START :

Initialize population with randomly selected possible solutions.

Find Fitness of each candidate solution in population.

REPEAT TILL (when **termination criteria** is satisfied):

Parent Selection Stage

Cross over the parents pairs

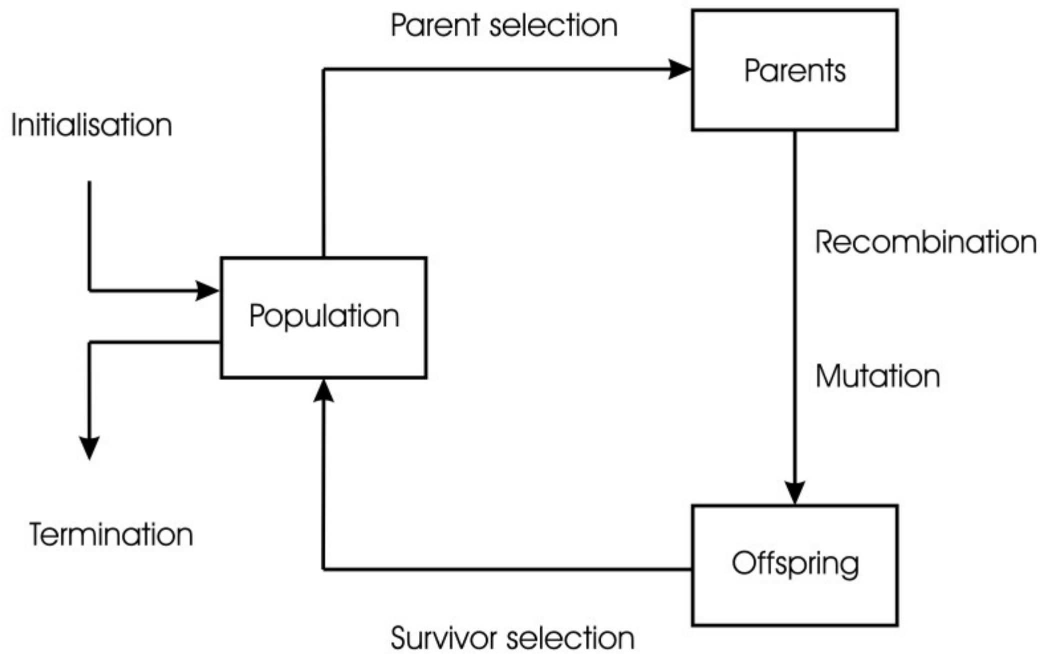
Mutate them

Find their fitness

Selection of next generation

LOOP ENDS

END



Here is pictorial representation of model.

Basic Components of Computational Model:

There are few important components of our presented model.

i. Representation (How we are representing our individuals ?)

ii. Fitness Function (That calculates fitness values of candidates)

iii. Population (Our Possible solutions)

iv. Parent Selection (That produces next generations)

v. Crossover and Mutation (Variety operators)

vi. Selection of survivors (Who will go to next generations?)

That was generic model based of Darwin's theory of evolution (Natural Selection).

We can represent individuals in many different way like binary string representation etc.

Fitness function plays a great role in evolutionary computation so it needs a bit care while selecting a fitness function.

Population is initialized randomly at start and constantly improve as we move on next generations.

Parents selection mechanism involves selection of parents that goes for crossover and mutation.

Mutation and crossover rates do matter a lot in evolutionary computation.

We must have a termination condition so that we can terminate. These conditions could be a particular threshold or maximum numbers of generations reached.

There are many other things like mutation and crossover rates, how much we have to mutate and what should be the optimal population size. These things are explained in application part with different experimentation results.

5. Application (Template Registration / Template Matching):

In computational model extraction part, basic generic model is presented on the basis of evolutionary theory and we called it evolutionary computational model. Now using this model we can solve optimization problems by tuning its basic components according to our needs.

In this report we focused on “Template matching problem” which is part of optimization problems domain and can be solved using evolutionary computational model.

We are given a small image (portion to be located) that we need to search from main large image (main template). Both images are gray scale images so that to ease its comparison.

Image is basically 2D grid of pixels and can be represented as 2D array / list.

Representation:

Now according to our presented model, first and most important part is representation of individuals in population. And each individual is a possible solution. In case of image it can be represented as a 2D point (x,y) coordinates. Each (x,y) point is a possible solution and can be a part of our initial population.

Fitness Function:

Fitness function used in this application evaluates each individual (compares it with small image) and return a value between -1 to 1. 1 means maximally matched and -1 means its inverse. Based on these values selection is done.

Population:

As mentioned above, population contains possible solutions and this application these are (x,y) points. Population size is changed in different trail tests.

Crossover:

In our application, crossover is done by selecting pair of parents, converting them into binary combined string and then on a random index their parts are swapped and two new individuals are created from 2 parents. Their binary representation is converted back to (x,y) points.

Mutation:

Mutation function takes two parameters, individual and its fitness value and according to some logical settings it mutates a random bit of individuals and returns its mutated version.

We have performed many experiments on changing population size, varying parent selection mechanism, survival selection mechanism and noted the results.

"Code file is also attached and code can be read from it."

6. Experiments:

Experiment No 1:

Kept population size large from 400 to 600 with high cross over rate and zero mutation rate".

In this experiment population size was large and every individual was involved in crossover but there was no mutation.

Approach was to generate children equal to number of parents. Then select best fit solutions from their multi set total equal to population size and discard rest unfit.

Result:

It seems to be more dependent on initial population . If initial population has some good possible solutions so it keeps moving around it.

I tested it with many rounds , only few times it founded 0.98 but maximally it stuck around 0.3 to 0.6

population size was = 500 and cross over rate 1.0 with 0 mutation. (cross over exploits I.e converges)

And only cross over is not working for me. As it stuck to optimal side. (not exploration).

Experiment No 2:

“With small population size of 50 and mutation and crossover rate 1 .”

In this approach I only selected top two fittest individuals and crossed over them. After mutation I added them population and then applied my survival selection mechanism and selected top 50 for next generation. But again selecting only top parents didn't worked. Very few times it reached 0.9 threshold but maximally it keeps on finding solutions near to parents and stuck in local search.

Experiment No 3:

With small population size equal to 50 with crossover and mutation rate 1 with no controlled mutation.

Fitness Approach: Generate 50 children from parents and then select best 50 among 100.

This thing show some increasing behavior but again it stuck in a local optima. More initial population oriented.

After experimenting with different approaches. Last but not least we tried a controlled smart mutation with same small population and crossover technique.

Experiment No 4:

After observing different experiments and trends in fitness values of each generation, we observed that by crossing over we are not going that far in our search space so we have to do mutation. But mutation (uncontrolled), caused problems as it provides us with wide jumps in search space and makes population diverse hence increases our search space. So in that experiment we defined some set of rules.

Once new individuals are created from parents, rather discarding whole set of parents or discarding half unfit from their combined multi set, we kept few top best parents in new generation and removed few worst off springs. Now mutation was controlled in that experiment.

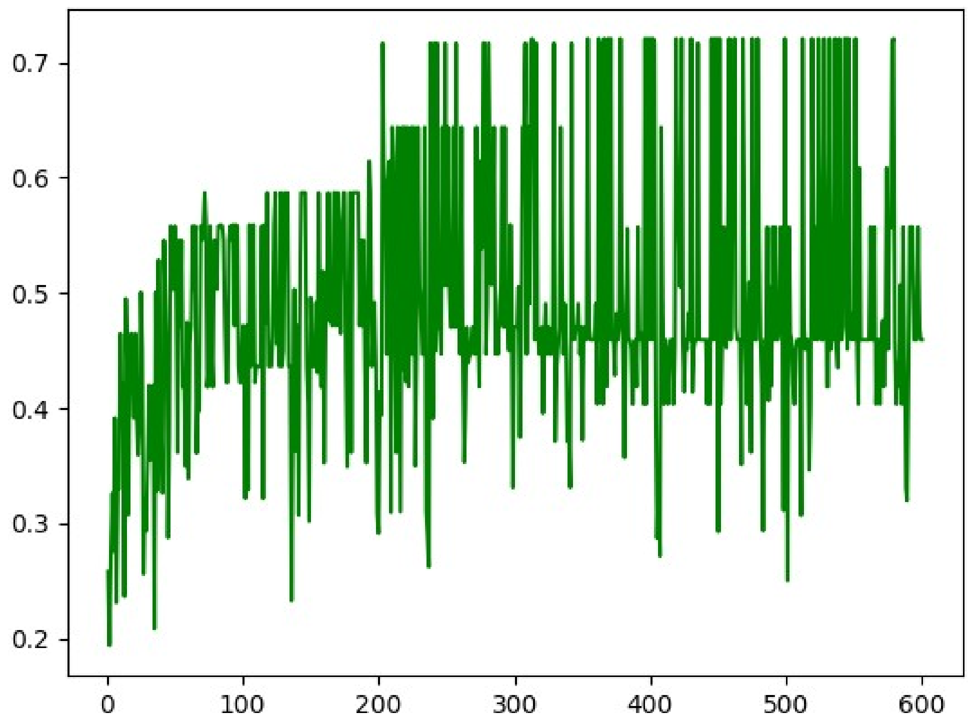
If fitness of our individual is below certain threshold we mutated it to great extent by flipping their most significant bits in binary representation. This thing provided us with big jumps in population and increased our searching space.

And one more close check was made , if our fitness value of individual is above a certain threshold like close to the destination, we reduced mutation by flipping only lower bits and this provided us with a very little change in off spring.

The results of this experiment was far far better than other approaches where mutation was random and uncontrolled. Clear increasing behavior was shown in result of this approach.

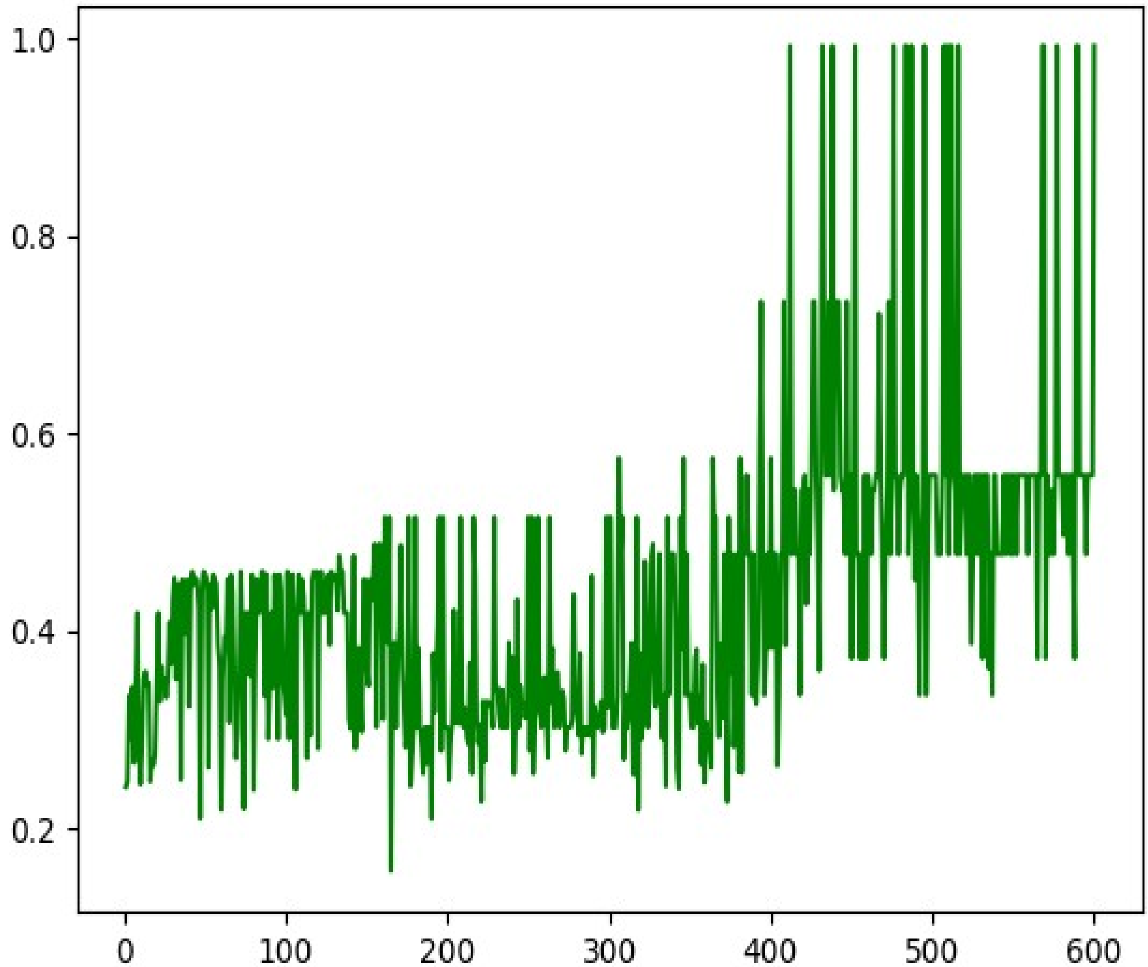
(You can see mutation function in code file for better understanding).

We have attached few graphs which represent best fitness values of each generation. On x axis are generations and on y axis are fitness values.



Graph No : 1

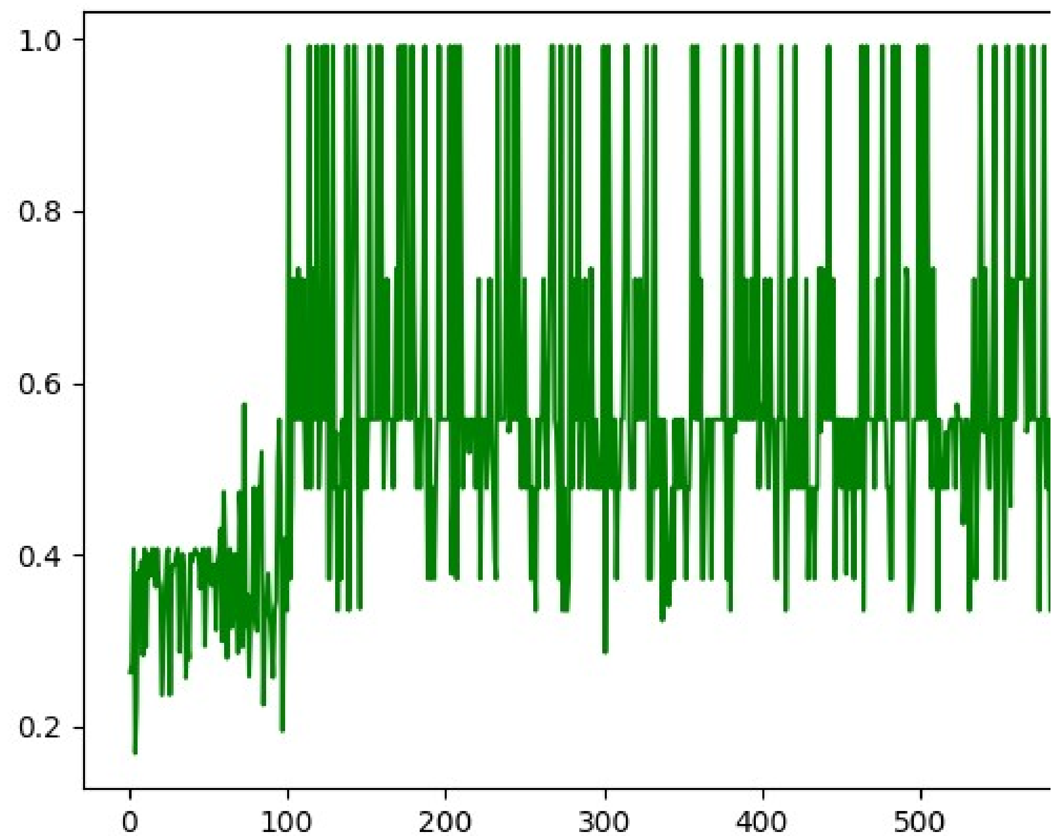
There are fluctuations in both graphs as we experimented with different mutation rates. In graph 2 , it is obvious we achieved 1.0 fitness values many times in total of 600 generations. And found baba jee face couple of times.



Graph No 2

Third graph is given below. Again clear increasing behavior can be observed.

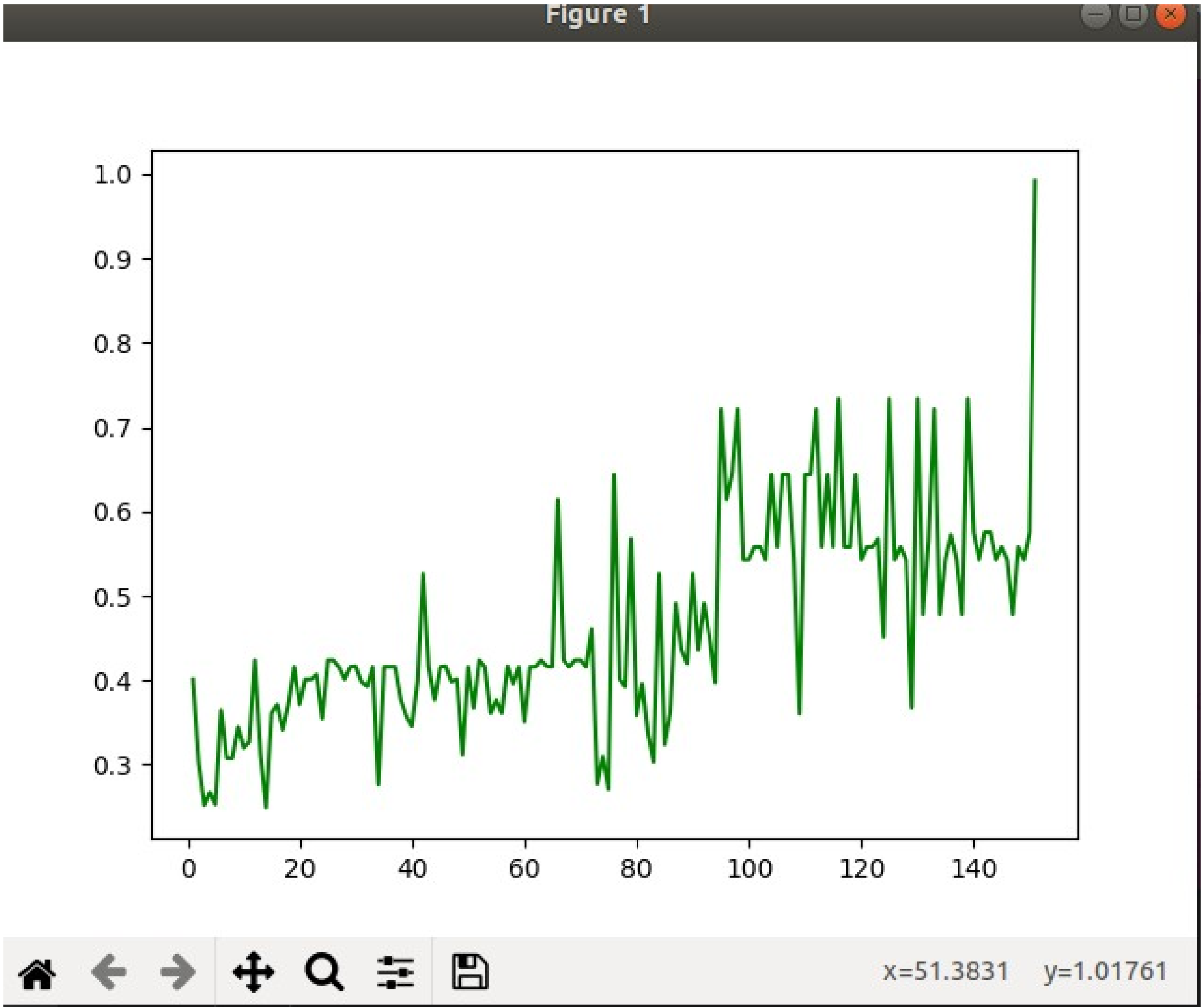
Pdf generated using libre office ubuntu



Graph No 3

Pdf generated using libre office ubuntu

Below is the final graph after setting good controlled mutations, here fluctuations are very less and within 150 generations we achieved 1.0 threshold and found baba jee accurately.



Pdf generated using libre office ubuntu

Green marker shows the location of baba jee template.

Experiment to be conducted:

After researching on internet sources we found that nowadays , dynamic parameter tuning techniques are being used in which probability for crossover and mutation are dynamically changing over time with each generation. And researchers mentioned that it improves overall performance of algorithm. Technique like “Dynamic Increasing of Low Mutation/Decreasing of High Crossover (ILM/DHC)” [Article : A review with new dynamic approach]. We will surely modify our algorithm according to these settings and will note the results.

7.Recommendations and Discussion:

As we shifted to controlled mutation because it was more correct according to basic understanding and its results were far far better as compared to unchecked mutation.

We do recommend to while selecting initial population you should be very careful, as if population is size is very low you can easily stuck in local optima while large population size become a random linear search. We also experimented with different population sizes are observed the behavior. Parent selection and next generation selection should be intelligent too, it is random but random with some intelligence.

Attached code is completely functioning you can simply run it through terminal by typing *python3 algorithm.py*

def main() is the main function in which other functions are called. There is separate function for each computation. Like for cross over there is function *def crossover(parent_one,parent_two)*

We haven't worked much for better graph and image locating visuals . So bear with that.

References :

[1] https://www.researchgate.net/post/How_to_calculate_the_Crossover_Mutation_rate_and_population_size_for_Genetic_algorithm

[2] https://en.wikipedia.org/wiki/Mutation_(genetic_algorithm)

[3] https://www.sciencedirect.com/topics/computer-science/mutation-probability

[4] https://biography.yourdictionary.com/articles/what-did-charles-darwin-invent.html

Pdf generated using libre office ubuntu

[5] Wikipedia contributors. "Charles Darwin." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 26 Oct. 2020. Web. 9 Nov. 2020.

“Report is generated using lib-re office ubuntu”